gene GIS: Computational Tools for Spatial Analyses of DNA Profiles with Associated Photo-Identification and Telemetry Records of Marine Mammals

C. Scott Baker
Oregon State University
Hatfield Marine Science Center
Newport, OR 97365-5296
phone: (541) 867-0255 fax: (541) 867 0138 email: scott.baker@oregonstate.edu

Dawn Wright
Department of Geosciences
Oregon State University
Phone: (541) 737-1229 fax: (541) 737-1200 email: dawn@science.oregonstate.edu

John Calambokidis
Cascadia Research
218 ½ W. 4th Ave.
Olympia, WA 98501
Phone: (360) 943-7325 ext 104 email: calambokidis@cascadiaresearch.org

Award Number: N000141110614
http://mmi.oregonstate.edu/ccgl

LONG-TERM GOALS

We will develop computation tools for improved visual exploration and spatial analysis of DNA profiles, with accompanying photo-identification records or telemetry tracks of marine mammals. Developments will include an integrated Geographic Information System (GIS) data model with an enhanced Graphical User Interface (GUI) for a stand-alone program within an ArcGIS framework and a web-based program in a more broadly accessible format for displaying individual identification photographs and information from linked DNA profiles. Referred to as *geneGIS*, the program will provide the ability to display, browse, select, filter and summarize spatial or temporal relationships of these individual-based records and associated datasets. A toolbox of software applications will allow basic summaries of spatially selected data and export of data in most standard formats (e.g., XLS, CSV, MDB, KML), as well as individual formats required for programs commonly used in genetic analyses of population differentiation, capture-recapture estimates of abundance, population assignment of individuals and estimates of kinship or parentage. The data formatting will comply with OBIS standards and the software architecture will be compatible with the Arc Marine model, providing a link with other datasets and tools needed for an integrated description of the genetic and environmental 'seascape' of cetaceans. We will implement *geneGIS* and the web-based application using DNA profiles and photo-identification records derived from an ocean-wide survey of humpback whales in the North Pacific (SPLASH), providing a comprehensive description of a complex migratory population. Such a description will be suitable for informing conceptual models of cetacean

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
30 SEP 2011		**00-00-2011 to 00-00-2011**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
gene GIS: Computational Tools for Spatial Analyses of DNA Profiles with Associated Photo-Identification and Telemetry Records of Marine Mammals	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Oregon State University,Hatfield Marine Science Center,2030 SE Marine Science Drive,Hatfield,OR,97365**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**7**	

populations, including the Population Consequences of Acoustic Disturbance (PCAD) and the Testing of Spatial Structure Methods (TOSSM).

OBJECTIVES

The overall objectives can be ordered into five tasks (with related subtasks):

- Task 1: Develop database architecture following Arc Marine model for integration and display of DNA profiles with photo-identification and telemetry records in a stand-alone ArcGIS framework, and enhance features of web-based application currently designed for display and visual exploration of photo-identification catalogues.
- Task 2: Develop ArcGIS tools for data query, visual exploration and basic statistical summaries for spatial and temporal partitions of individual-based records (DNA profiles, photo-identification records and telemetry tracks).
- Task 3: Enhance user-directed spatial/temporal selection and export of individual-based records for advanced statistical analyses. This will include tools to export data compatible with existing software used for genetic analyses and capture-mark-recapture.
- Task 4: Demonstrate functionality of *geneGIS* and web-based application through importation and integration of existing large-scale datasets of DNA profiles and photo-identification records from the Structure of Populations, Levels of Abundance and Status of Humpbacks program in the North Pacific (SPLASH).
- Task 5: Prepare a comprehensive user guide for all software functions and analyses implemented in the system.

APPROACH

The computation development of geneGIS is pursuing two approaches: 1) as a stand-alone program for displaying individual identification photographs and information from linked DNA profiles within an ArcGIS framework; and 2) as a web-based program in a more broadly accessible format. The intent is to benefit from the strengths of each approach while assuring compatibiliy and interoperability through a common database architecture.

The ArcGIS approach is being directed by the PI through Oregon State University, with support from Prof Dawn Wright (on leave with ESRI), her PhD student, Dori Dick, and members of the Marine Mammal Institute, including Dr. Beth Slikas, Tomas Follet and Debbie Steel. This approach takes advantage of previous experience with management of whale telemetry database under the Arc Marine model (Lord-Castillo *et al.* 2009; Wright *et al.* 2007). The web-based approach is being developed under subcontract to Jason Holmberg of the Shepherd Project, with support of John Calambokidis and Erin Falcone of Cascadia Research Cooperative. This approach takes advantage of an existing open-source software framework supporting capture-mark-recapture (CMR) studies of marine megafauna by the Shepherd Project (Holmberg *et al.* 2008). This software framework provides a scalable, Web-based platform for CMR data management (http://www.ecoceanusa.org/shepherd/doku.php?id=start) and was selected by Cascadia Research to develop and host the SPLASH Photo-ID Catalog (available in beta version as http://www.splashcatalog.org). With support from the *geneGIS* initiative, the beta version of

the SPLASH Photo-ID Catalog is being enhanced to include genetic data (haplotypes and microsatellite markers), allowing for the reconciliation of genotype and photo-identification catalogs.

WORK COMPLETED

Since the contract was awarded in mid-year, we have focused on development of a joint database architecture to relate the individual-based records to the Arc Marine standards and conventions of the Shepherd Project, as described in Task 1. The reconciliation of the SPLASH photo-identification records and available DNA profiles is underway through integration and crosschecking by Cascadia and MMI. An initial reconcilliation of photo-identification records and genetic records from SPLASH was completed to provide a template for the database architecture and to facilitate data import/export modules. This involved a subset of the total SPLASH catalogue and genetic dataset, representing California/Oregon feeding region and the Central American breeding region. These are two of the smaller SPLASH study regions with a fairly high exchange rate based on the photographic dataset, and thus provide a good test case for data integration.

RESULTS

Database architecture. A provisional architecture has been agreed to accommodate relational databases typical of those used in the collection of individual-based records from photo-identification, telemetry and the collection of tissue samples for genetic analyses and ecomarkers. The architecture and nomenclature conform to Arc Marine and Darwin Core standards where possible and can accommodate the current databases developed for telemetry data at MMI and SPLASH collection records at Cascadia Research.

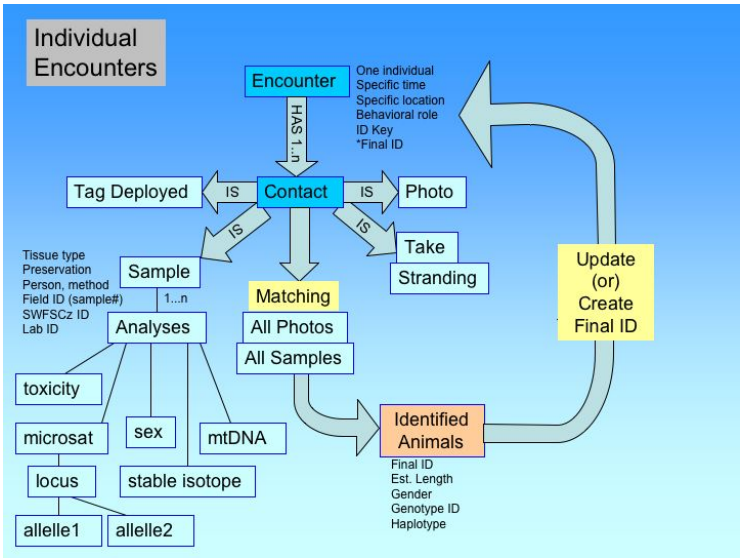

Figure 1: Database architecture for geneGIS. The architecture asssumes an individual 'encounter' during which there is the collection one of more datatypes or deployments of instrument. These 'contacts' are related to each other by an encounter-specific 'ID key' (see also Figure 3). Matching of photo-IDs or genotypes establish links between encounters through a 'Final ID' code, similar to that used in the current SPLASH photo-ID catalogue. A joint 'Final ID' may be dynamically allocated where there are pre-existing catalogues of photo-IDs or genotypes.

Reconciled SPLASH datasets. A preliminary integration (reconciliation) of microsatellite genotype identifications and photo-identification was conducted using data collected at breeding areas off Central America and feeding areas off California and Oregon (CA-OR) during the SPLASH project. There were 604 unique individuals photo-identified in the two regions: 525 from CA-OR, 105 from Central America, with 26 photographed in both regions. Individual whales were photographed on from 1 to 19 different encounters during the study. A total of 182 tissue samples were collected from these regions during the study, 124 of which included a corresponding collection of a photo-ID. Analysis of these samples, including sex, mtDNA haplotypes and 10 microsatellite loci, yielded 164 unique genotypes, with 10 whales sampled on two or three occasions and subsequently matched genetically.

The overlap between the photographic and genetic identification collections provides several opportunities to refine and extend the two data types across the larger sample. In most studies of this type, many more photographs are collected than tissue samples; however, when a tissue sample is linked to a photo-ID through contacts in at least one encounter (see Figure 1), and that individual is photographed on another occasion but not sampled, we can 'extend' the single source of genetic information to any additional photo-documented occurrences. In this way, we were able to extend the genetic data, including sex, mtDNA haplotype, and genetic identity, to 18 other occurrences when a whale was photographed but not sampled in these two regions during SPLASH. This is particularly valuable when a whale is sampled and photographed in one region, but only photographed in another, as happened with three individuals in this small sample (see Figure 2). We expect this type of extension of genetic data across photographic histories to become much more common when the genetic data is integrated across the larger SPLASH dataset of more than 18,000 photo-ID encounters and 2,100 genotype encounters.

ArcGIS query and display. A toolbox for data query, visual exploration and connectivity for spatial and temporal partitions of individual-based records is under development in ArcGIS vs10. The reconciled dataset of SPLASH photo-ID and genotype records from California/Oregon and Central America (see above), with associated geo-coordinates, has been imported into ArcGIS as the basis for visual display and exploration (Figure 2).

Web-based integration. The Web-based platform of the Shepherd Project and the existing SPLASH Photo-ID Catalog are being modified to incorporate genetic data (sex, mtDNA haplotypes and microsatellite markers), allowing for the integration of photo-identification and genotype catalogs into a cohesive framework for Capture-Mark-Recapture (CMR). By developing this functionality in the context of the Shepherd Project, *geneGIS* will ensure that the tools and techniques developed in this study are broadly applicable to CMR across other species and projects.

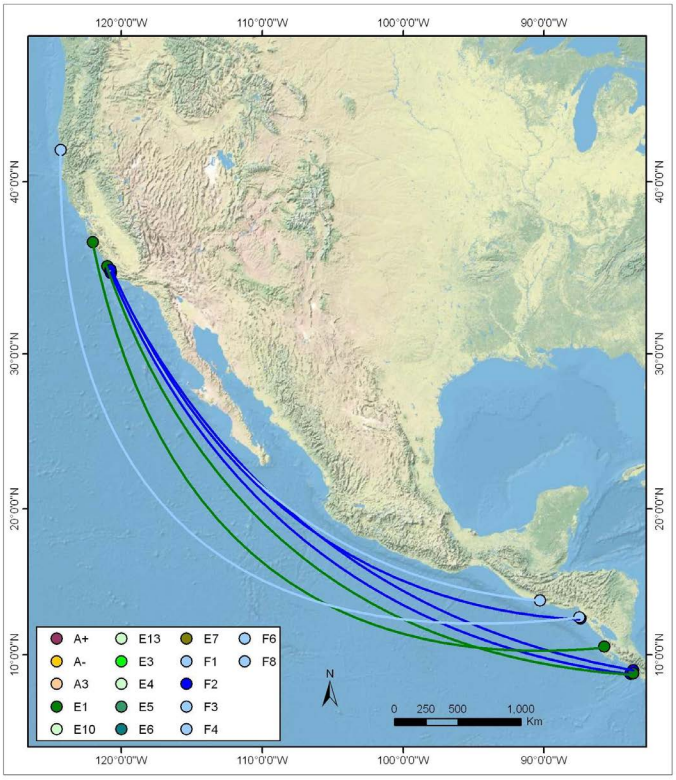

Figure 2: The location and connectivity of individual-based encounters for humpback whales photo-identified in both California and Central America, with a corresponding genetic sample in one or both region, during the SPLASH project. The figure is the result of a four-step process using ArcGIS query and display tools: 1. Query the database for all records of matching SPLASH ID records between California and Central America. 2. Filter for matching records with at least one corresponding genetic record (GeneID); 3. Display the records on a map 'painted' or symbolized according to mtDNA haplotype; and 4. Connect the pairs of matching records with a line. The key shows symbol colors assigned to the 17 mtDNA haplotypes found in the two regional datasets. The lines are not intended to represent routes of individual movement, only to connect the locations of regional encounters of the 7 individuals.

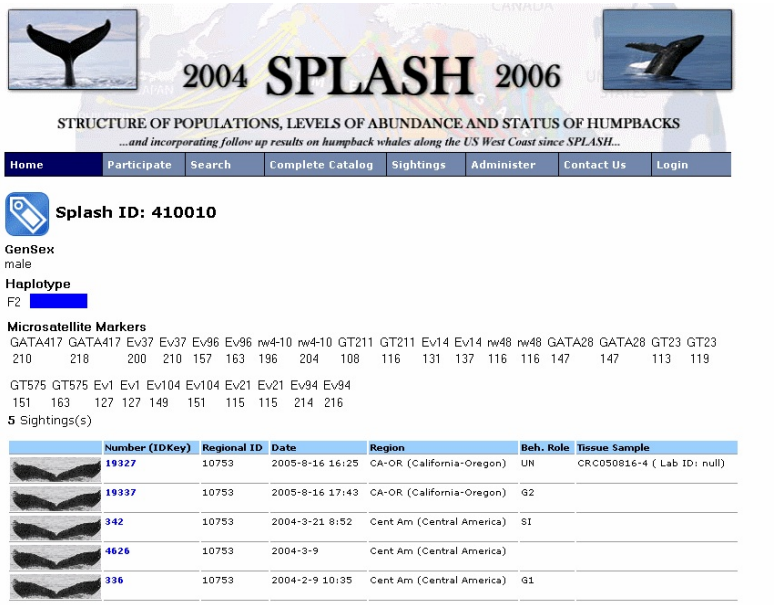

Figure 3: The provisional display of information for a single 'Final ID' consisting of five encounters (IDkeys), one of which included the collection of a genetic sample, in the beta version of SPLASH Photo-ID Catalog. The genetic sample, collected during an encounter in Central California, can be extended to other photo-ID encounters, including those in Central America (see Figure 2). These representative records were selected from the reconciled dataset for photo-identification and genetic records from the California/Oregon and Central America regions of SPLASH.

IMPACT/APPLICATIONS

A growing number of large-scale studies of marine mammals and other marine megafauna (e.g., sharks, and turtles) are collecting spatially explicit records linked through individual identification to genetic samples, photo-identification and telemetry. These spatio-temporal records have been used to track the migration and life history parameters of individuals, to estimate the abundance and trends of populations and, in the case of genetic markers, to infer close kinship (e.g., parent/offspring relationships) and define management units, or Distinct Population Segments. To date, however, there has been a conspicuous absence of computational tools for integration and spatial exploration of these individual records, particularly the potential for linking photo-identification to genetic information (e.g., DNA profiles) and for extending genetic identity to include close kinship. We anticipate that *geneGIS* will help to fill this gap between available datasets and computation tools, improving our understanding of cetacean populations and human impact on these populations.

RELATED PROJECTS

Title: '*Examination of health effects and long-term impacts of deployments of multiple tag types on blue, humpback, and gray whales in the eastern North Pacific*' with funding Cascadia Research Collective, from the National Oceanographic Partnership Program (NOPP) and Interagency Committee

on Ocean Science and Resource Management Integration (ICOSRMI). In collaboration with Cascadia Research, the Marine Mammal Institute (MMI), Oregon State University (OSU) is assisting with the integration of photo-identification records and associated genetic samples, to improve understanding of long-term impact of satellite tagging. The resulting database should be suitable for implementation in *geneGIS*.

Title: '*The Shepherd Project'*. This project started as a collaborative software platform for globally coordinated whale shark research, as described in the http://www.ecoceanusa.org/. The success of this platform in managing and supporting the growth of the whale shark catalog led to its selection for the Web-based implementation of the SPLASH Photo-ID Catalog (http://www.splashcatalog.org). Through ongoing development of this open-source platform, the Shepherd Project provided for the cross application of new functionality to other long-term studies of individually identified marine mammals or marine megafauna.

REFERENCES

Holmberg, J., B. Norman and Z. Arzoumanian. 2008. Robust, comparable population metrics through collaborative photo-monitoring of whale sharks *Rhincodon typus*. Ecological Applications 18:222-233.
Lord-Castillo, B.K., B.R. Mate, D.J. Wright and T. Follett. 2009. A Customization of the Arc Marine Data Model to Support Whale Tracking via Satellite Telemetry. Transactions in GIS 13:63–83.
Wright, D.J., M.J. Blongewicz, P.N. Halpin and J. Breman. 2007. Arc marine: GIS for a blue planet. ESRI, Inc.

HONORS/AWARDS/PRIZES

The PI, C. Scott Baker, Oregon State University, was awarded a Pew Fellowship in Marine Conservation for 2011-2013 to support a study of seascape genetics in dolphins of Oceania. The project, refered to as 'A Pattern of Dolphins', is expected to have strong synergies with the *geneGIS* program.